BMJ

BMJ 2012;344:e686 doi: 10.1136/bmj.e686 (Published 21 February 2012)

RESEARCH METHODS & REPORTING

Assessing the value of diagnostic tests: a framework for designing and evaluating trials

The value of a diagnostic test is not simply measured by its accuracy, but depends on how it affects patient health. This article presents a framework for the design and interpretation of studies that evaluate the health consequences of new diagnostic tests

Lavinia Ferrante di Ruffano *research fellow*[1], Christopher J Hyde *professor of public health and clinical epidemiology*[2], Kirsten J McCaffery *associate professor and principal research fellow*[3], Patrick M M Bossuyt *professor of clinical epidemiology*[4], Jonathan J Deeks *professor of biostatistics*[1]

[1]Department of Public Health, Epidemiology, and Biostatistics, School of Health and Population Sciences, University of Birmingham, Birmingham B15 2TT, UK; [2]PenTAG, Institute for Health Services Research, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK; [3]Screening and Test Evaluation Program, School of Public Health, University of Sydney, Sydney, Australia; [4]Department of Clinical Epidemiology and Biostatistics, Academic Medical Centre, University of Amsterdam, Amsterdam, Netherlands

Most studies of diagnostic tests evaluate only their accuracy. Although such studies describe how well tests identify patients with disease (sensitivity) or without disease (specificity), further evidence is needed to determine a test's true clinical value. Firstly, since tests are rarely used in isolation, studies are needed that evaluate the performance of testing strategies, accounting for when and how a new test is used within a diagnostic pathway, and how its findings are combined with results of other tests.[1] Secondly, decision making involves selecting among multiple testing strategies; thus studies that compare test strategies and estimate differences in sensitivity and specificity are more informative than those that evaluate the accuracy of one test or diagnostic strategy.[2] Thirdly, improvements in test accuracy will not benefit patients unless they lead to changes in diagnoses and patient management, requiring evaluations of the effect of improved accuracy on decision making.[3] Finally, improved decision making is only one route by which tests affect patient health, and empirical evaluations are needed to compare the effect of test strategies on patient health.[4]

Ideally, new tests should only be introduced into clinical practice if evidence indicates that they have a better chance of improving patient health than existing tests.[5 6] Tests can be compared by evaluating the downstream consequences of testing on patient outcomes, either directly in a randomised controlled trial or by decision analysis models that integrate multiple sources of evidence. Test-treatment trials randomly allocate patients to tests, follow up subsequent management, and measure outcomes only after treatment has been received (fig 1⇓).[7] Decision models use existing clinical data to extrapolate, through a number of assumptions, the link between intermediate outcomes (such as accuracy) and long term outcomes.[8] A key issue for trials and decision models is the selection of outcomes that need to be measured or modelled to evaluate how tests are affecting patients. This selection requires a priori knowledge of the mechanisms by which tests affect patient health.

In this article, we provide a comprehensive review of the mechanisms that can drive changes to patient health from testing, and include a summary checklist to assist readers, researchers, and funders who wish to design or appraise studies evaluating diagnostic tests. We have based our framework on a review of a large cohort of published test-treatment trials [9] and key methodological literature.

Effect of tests on patient health

To establish whether a new diagnostic test will change health outcomes, it must be examined as part of a broader management strategy. Testing represents the first step of a test-treatment process: (1) a test is administered to identify a target condition, (2) the test result is considered (3) alongside other evidence to decide a diagnosis, and (4) a course of treatment is identified (5) and implemented (fig 2⇓).[10]

Changes to any aspect of this pathway after the introduction of a new test could trigger changes in health outcomes. Table 1⇓ lists the mechanisms that commonly affect health outcomes.

Direct test effects

Test procedure

Some diagnostic procedures carry a risk of harm, hence alternatives that offer reduced procedural morbidity will be of immediate benefit to patients. For example, use of sentinel lymph node biopsy rather than dissection of the axillary node

Correspondence to: Jonathan J Deeks j.deeks@bham.ac.uk

to investigate metastatic spread in patients with early breast cancer results in much lower rates of postoperative swelling of the arm, seroma formation, numbness, and paraesthesia.[11]

Altering clinical decisions and actions

Feasibility and interpretability

The downstream value of a test will be impaired at the outset if there are contraindications to its use or if it is prone to technical failure (feasibility), while tests that are more difficult to interpret (interpretability) could produce fewer definitive results. Either problem could require additional investigations, increasing the time to diagnosis, or reducing diagnostic and therapeutic yields through incorrect decision making or poor diagnostic confidence.

We observed this in a trial evaluating the diagnosis of coronary artery disease. Patients with acute chest pain who were allocated to exercise electrocardiography were significantly more likely to be referred for further investigation (coronary angiography) than those allocated to stress echocardiography.[12] This finding was caused by the higher frequency of inconclusive diagnoses produced by exercise electrocardiography, some of which were because the test was contraindicated.

Test accuracy, diagnostic yield, therapeutic yield, and treatment efficacy

More accurate tests will improve patient outcomes if the reductions in false positive or false negative results lead to more people receiving appropriate diagnoses (diagnostic yield) and appropriate treatment (therapeutic yield). The degree to which appropriate treatment can improve patient outcomes depends on its efficacy (treatment efficacy). In a trial evaluating the effect of fluorescence cystoscopy on the recurrence of bladder carcinoma in situ, the enhanced accuracy of fluorescence cystoscopy compared with white light cystoscopy alone led to a substantial increase in lesions being identified and treated at initial diagnosis, which significantly reduced the rate of recurrence.[13]

Diagnostic and therapeutic confidence

Although diagnostic yield generally increases with accuracy, it is also affected by a doctor's confidence in the diagnostic test. Tests inducing greater confidence could benefit patients by reducing the need for further investigations and shortening the time to treatment. The results of a trial evaluating the triage of patients with non-small cell lung cancer who were referred for operative staging with positron emission tomography (PET), show how a lack of diagnostic confidence can over-ride the benefits of improved accuracy.[14] PET identified patients for whom surgery was not indicated because of incurable mediastinal disease, but no difference was found in the proportion of patients avoiding a thoracotomy (the primary outcome) because surgeons still preferred to confirm PET findings using standard operative staging.

Doctors' confidence in the ensuing success of a treatment plan can affect treatment effectiveness by influencing the approach to treatment, particularly in surgery. Digital subtraction angiography (DSA) and multidetector row computed tomographic angiography (MDR-CTA) can both determine the location and degree of vascular narrowing in patients with symptomatic hardening of peripheral arteries. Doctors using DSA were significantly more confident of plans for surgery, owing to the test's clearer vascular images; however, MDR-CTA images were found to obscure interpretation and decrease confidence in the presence of vessel wall calcifications.[15]

Changing timeframes of decisions and actions

Tests that are undertaken earlier or produce results more quickly can improve health outcomes. For example, patients with unstable angina and non-ST segment elevated myocardial infarction allocated to receive early coronary angiography had a reduced risk of death, non-fatal cardiac events, and readmission.[16] Patients with ventilator associated pneumonia allocated to a rapid antimicrobial susceptibility test received definitive results on average 2.8 days earlier than those receiving the standard susceptibility test and experienced significantly fewer days of fever, bouts of diarrhoea, and days on mechanical ventilation.[17]

However, quicker results are beneficial only if they produce earlier diagnosis or treatment. The addition of polymerase chain reaction (PCR) to conventional analysis of nasopharyngeal swabs for distinguishing between viral and bacterial causes of lower respiratory tract infection failed to decrease time to treatment, because physicians were unwilling to base treatment decisions solely on PCR, preferring to wait for slower bacterial results.[18] Earlier diagnosis can provide psychological benefit by dispelling anxiety or providing earlier reassurance but can also cause psychological harm, particularly if effective treatments are unavailable. The psychosocial impacts of an earlier diagnosis have been highlighted in women following a positive cervical smear test[19] or mammogram.[20]

Influencing patient and clinician perceptions

The patient's perspective and the doctor's personal perspective can also influence decision making, sometimes in unexpected ways. These unpredictable responses can eliminate or enhance potential improvements gained from other aspects of the test-treatment pathway.

Patients

Patients' perceptions of testing, their experience of the testing process, and their understanding of the test result can all affect downstream health. Many studies show social, emotional, cognitive, and behavioural effects of testing across various clinical conditions.[21]

Test-treatment pathways will be unsuccessful if patients are unwilling to undergo a procedure. This is especially important if multiple testing is required; an unpleasant first test can adversely influence patients' willingness to attend follow-up testing or treatment. The experience of undergoing tests can also influence illness beliefs. In a randomised trial, women who were able to observe their diagnostic hysteroscopy on a screen were reportedly less optimistic about the effectiveness of treatment offered, experienced more anxiety, but were better able to deal with procedural discomfort than women who could not see the screen.[22]

Diagnostic placebo effects might occur if the impression of a thorough investigation improves perceptions of health status. This could account for the significant improvements in health utility that were reported by patients with acute undifferentiated chest pain diagnosed in a specialist unit, compared with those diagnosed in emergency departments, despite having equivalent treatment and rates of adverse cardiac events.[23]

Receiving a diagnosis can have behavioural and health consequences—for example, by confirming patients' negative

health beliefs. Patients with lower back pain reported higher pain scores and poorer health status after receiving an x ray than those who received only a standard consultation.[24] The incidental diagnosis of non-pathological abnormalities may have given patients a reason for their pain and encouraged illness behaviour despite the absence of an organic cause.

Adherence to treatment

Patients' experiences and perceptions of the test-treatment pathway will also affect downstream health behaviours, such as the willingness or motivation to adhere to medical advice.[25] Negative perceptions or experiences of testing and clinical diagnosis could cause patients to lose confidence in the diagnosis or management plan, making them reluctant to have subsequent testing or treatment.

Doctors

Doctors' emotional, cognitive, social, or behavioural perspectives, although external to objective medical concerns, are nevertheless important in decision making. Referring doctors might modify management to reassure and satisfy patients or to prevent perceived threats of malpractice, often by requesting additional diagnostic information.[26] This defensive medicine tends to raise the diagnostic threshold needed to trigger a change in management,[27] and if additional tests are less accurate, harmful, or lead to treatment delays, patients will be adversely affected.

Systemic approach to evaluating tests

These examples establish that diagnostic tests often affect patient health outcomes in many complex ways. Although test accuracy is commonly regarded as the main mechanism to influence clinical effectiveness,[28] we caution against its use as a surrogate for patient health. Only by looking at the test-treatment pathway as a whole can we identify which outcomes need to be evaluated to fully capture a test's health effects.

Sound evaluations of healthcare demand explanation of how the intervention will improve patient health.[29] This is equally true of diagnostic tests, although they are considerably more challenging to evaluate because so many intermediate, interacting factors are at stake. The need to identify which of these factors will exert an effect and how, is a key tenet of complex intervention guidance. [30] Table 2⇓ provides a list of questions to guide the structured assessment of which processes are relevant and need to be measured within a given diagnostic comparison. This approach highlights precisely where in a test-treatment pathway important differences might originate, and will be useful for designing studies, appraising existing research, and determining what new evidence is needed to formulate diagnostic guidelines (box).

We identify three benefits from using this framework. Firstly, it presents a structure for carefully developing a rationale that underpins the performance of a putative testing strategy. Secondly, it guides the identification of outcomes for randomised controlled trials, and will also assist in constructing appropriate decision models, particularly when trials are not practicable.[32] Finally, the approach supports a full interpretation of empirical results by enabling trialists to distinguish between true ineffectiveness, poor protocol implementation, and methodological flaws in the study design.[33] These tasks are particularly important for trials of tests, where sample sizes often need to be several orders of magnitude larger than they do in trials of treatments to detect differences in patient outcomes (fig 3⇓).[34] Findings of no effect are all too often interpreted as "evidence of absence," when in reality studies rarely make provision for being able to attribute negative results to the diagnostic intervention, the study design, or (importantly) an inconsistently implemented test-treatment strategy. These interpretations can be distinguished by identifying and measuring the relevant driving mechanisms. By recording the use of additional diagnostic tests, treatments, and decision making, the failure of PET to reduce the rate of thoracotomies in patients with non-small cell lung cancer was shown to lie with an ill conceived treatment strategy, rather than with efficacy of the test.[14] The trialists identified patients for whom PET, unexpectedly failed to change management decisions, and they then found that strong preferences for the existing management (to operate on all patients with stage IIIa disease) exceeded the effect of PET results. By identifying all relevant mechanisms, and measuring how they exert their effect, test-treatment trials are more likely to contribute important evidence to the use of tests in clinical practice.

Conclusion

Establishing benefit to patient health must be the priority for diagnostic evaluations. Test accuracy is one component of test evaluation, but does not capture the impact of tests on patients. By considering the ways in which tests affect patients' health, we reiterate the complex intervention perspective[30] that it is not sufficient to measure outcomes, but rather it is essential to understand how these outputs are created, by conducting analyses of their workings and the mechanisms that underpin them. Clearly, this process must be undertaken with expert and stakeholder consultation to ensure all influential mechanisms are identified.

Contributors: JJD conceived the idea for this project with support from CJH. LFR did most of the primary research for the paper. The initial framework was devised by LFR, CJH, and JJD, and further refined by all authors. All the authors drafted, revised, and gave final approval to the article. JJD is the guarantor.

Funding: The development of the framework was funded partly by the UK Medical Research Council Methodology Programme (grant G0600545, awarded to JJD), as part of a wider investigation into the use of randomised trials for evaluating the clinical effectiveness of diagnostic tests. The funders had no involvement in the research project. JJD is partly supported by the Medical Research Council Midland Hub for Trials Methodology Research, University of Birmingham (grant G0800808).

Competing interests: All authors have completed the ICJME unified disclosure form at www.icmje.org/coi_disclosure.pdf and declare: the work was funded partly by the UK Medical Research Council Methodology Programme; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006;332:1089-92.
2 Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM, on behalf of the Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 2008;149:889-97.
3 Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88-94.
4 Lijmer JG, Bossuyt PMM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol* 2009;62:364-73.
5 Fineberg HV. Evaluation of computed tomography: achievement and challenge. *Am J Roentgenol* 1978;131:1-4.

Box: Example evaluation of a diagnostic test

Consider replacing conventional imaging (which usually involves multiple images with different technologies) with PET-computed tomography (CT) for the diagnosis of breast cancer recurrence in adults with clinically suspected tumours. The first step is to state the alternative diagnostic and management pathways that will be compared, and to note the differences between them to narrow down mechanisms to consider.

How PET-CT improves patient health

On the basis of a recent systematic review, we might expect the improved accuracy of PET-CT to be the main mechanism driving changes to health.[31] A more accurate differentiation of patients with and without recurrence would increase diagnostic and therapeutic yields, and the treatment consequences thereof.

Accuracy improvements could be offset by other decisions; patient contraindications to PET-CT could mean patients must revert to the existing multitest strategy. Although the known technical capabilities of PET-CT might increase doctors' confidence, the obligation to rely on the results of one test could initially weaken such confidence, thus reducing the effective accuracy of the new protocol.

By contrast, use of a single test could accelerate treatment by enabling a quicker diagnosis. Nevertheless, the requirement for a specialist to interpret PET-CT scans could mitigate this benefit. Comparative procedural harms might also differ, highlighting the importance of considering direct health outcomes, although conventional imaging usually requires CT, so the exposure to radiation as a consequence of using PET-CT is probably similar. However, the success with which the new strategy operates will depend on any differences in perceptions and experiences; PET-CT might be more or less reassuring to patients and clinicians, and these unknown influences would need to be measured carefully.

Choosing outcomes to evaluate PET-CT

Using the framework prompts the consideration of informative outcomes by showing the new pathway's full range of health effects, and allowing the assessment of all relevant direct and downstream measures of important patient outcomes. In the present example, such outcomes might include measures of anxiety, reassurance, health beliefs, function, symptoms, recurrence, progression, and survival.

Identified mechanisms can be measured as process outcomes in order to assess whether the new pathway is operating as expected. For example, the impact of temporality could be assessed as the time to diagnosis or time to treatment, and diagnostic confidence might be measured directly or by the number of additional investigations ordered.

Summary points

- The value of diagnostic tests ultimately lies in their effect on patient outcomes
- Tests can affect patient health by changing diagnostic and treatment decisions, affecting time to treatment, modifying patient perceptions and behaviour, or putting patients at risk of direct harm
- Improved accuracy is not always a necessary prerequisite for improving patient health, nor does it guarantee other downstream improvements
- All elements of the management process (including decision making and treatment) must be considered when evaluating a diagnostic test
- Randomised controlled trials of tests can measure these processes directly to understand why and how changes to patient health have occurred

6 Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;336:1106-10.
7 Canadian Critical Care Trials Group. Randomized trial of diagnostic techniques for ventilator-associated pneumonia. *N Engl J Med* 2006;355:2619-30.
8 Sutton AJ, Cooper NJ, Goodacre S, Stevenson M.Integration of meta-analysis and economic decision modeling for evaluating diagnostic tests. *Med Decis Making* 2008;28:650-67.
9 Ferrante di Ruffano L, Davenport C, Eisinga A, Hyde C, Deeks JJ. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. *J Clin Epidemiol* 2012;65:282-7.
10 Bossuyt PMM, Lijmer JG. Traditional health outcomes in the evaluation of diagnostic tests. *Acad Radiol* 1999;6:S77-80.
11 Purushotham AD, Upponi S, Klevesath MB, Bobrow L, Millar K, Myles J, et al. Morbidity after sentinel lymph node biopsy in primary breast cancer: results from a randomized controlled trial. *J Clin Oncol* 2005;23:4312-21.
12 Jeetley P, Burden L, Senior R. Stress echocardiography is superior to exercise ECG in the risk stratification of patients presenting with acute chest pain with negative Troponin. *Eur J Echocardiogr* 2006;7:155-64.
13 Babjuk M, Soukup V, Petrík R, Jirsa M, Dvorácek J. 5-aminolaevulinic acid-induced fluorescence cystoscopy during transurethral resection reduces the risk of recurrence in stage Ta/T1 bladder cancer. *BJU Int* 2005;96:798-802.
14 Viney RC, Boyer MJ, King MT, Kenny PM, Pollicino CA, Mclean JM, et al. Randomized controlled trial of the role of positron emission tomography in the management of stage I and II non-small-cell lung cancer. *J Clin Oncol* 2004;22:2357-62.
15 Kock MC, Adriaensen ME, Pattynama PM, van Sambeek MR, van Urk H, Stijnen T, et al. DSA versus multi-detector row CT angiography in peripheral arterial disease: randomized controlled trial. *Radiology* 2005;237:727-37.
16 Cannon CP, Weintraub WS, Demopoulos LA, Vicari R, Frey MJ, Lakkis N, et al. Comparison of early invasive and conservative strategies in patients with unstable coronary syndromes treated with the glycoprotein IIb/IIIa inhibitor tirofiban. *N Engl J Med* 2001;344:1879-87.
17 Bouza E, Torres MV, Radice C, Cercenado E, de Diego R; Sánchez-Carrillo C, et al. Direct E-test (AB Biodisk) of respiratory samples improves antimicrobial use in ventilator-associated pneumonia. *Clin Infect Dis* 2007;44:382-7.
18 Oosterheert JJ, van Loon AM, Schuurman R, Hoepelman AI, Hak E, Thijsen S, et al. Impact of rapid detection of viral and atypical bacterial pathogens by real-time polymerase chain reaction for patients with lower respiratory tract infection. *Clin Infect Dis* 2005;41:1438-44.
19 McCaffery KJ, Irwig L, Turner R, Chan SF, Macaskill P, Lewicka M, et al. Psychosocial outcomes of three triage methods for the management of borderline abnormal cervical smears: an open randomised trial. *BMJ* 2010;340:b4491.
20 Barton MB, Morley DS, Moore S, Allen JD, Kleinman KP, Emmons KM et al. Decreasing women's anxieties after abnormal mammograms: a controlled trial. *J Natl Cancer Inst* 2004;96:529-38.
21 Bossuyt PMM, McCaffery K. Multiple pathways and additional patient outcomes in evaluations of testing. *Med Decis Making* 2009;29:E30-8.
22 Ogden J, Heinrich M, Potter C, Kent A, Jones S. The impact of viewing a hysteroscopy on a screen on the patient's experience: a randomised trial. *BJOG* 2009;116:286-93.
23 Goodacre SW, Nicholl J, Dixon S, Cross E, Angelini K, Arnold J, et al. Randomised controlled trial and economic evaluation of a chest pain observation unit compared with routine care. *BMJ* 2004;328:254-60.
24 Djais N, Kalim H. The role of lumbar spine radiography in the outcomes of patients with simple acute low back pain. *APLAR J Rheumatol* 2005;8:45-50.
25 Haynes RB, Ackloo E, Sahota N, McDonald HP, Yao X. Interventions for enhancing medication adherence. *Cochrane Database Syst Rev* 2008;2:CD000011.
26 Summerton, N. Positive and negative factors in defensive medicine: a questionnaire study of general practitioners. *BMJ* 1995;310:27-9.
27 Hauser MJ, Commons ML, Bursztajn HJ, Gutheil TG. Fear of malpractice liability and its role in clinical decision making. In: Gutheil TG, Bursztajn HJ, Brodsky A, Alexander V, eds. Decision making in psychiatry and the law. 1st ed. Williams and Wilkins, 1991.
28 Hunink MGM, Krestin GP. Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology* . 2002;222:604-14.
29 Moher D, Hopewell S, Schultz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
30 Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: new guidance. Medical Research Council, 2008.
31 Pennant M, Takwoingi Y, Pennant L, Davenport C, Fry-Smith A, Eisinga A, et al. A systematic review of positron emission tomography (PET) and positron emission tomography/computed tomography (PET/CT) for the diagnosis of breast cancer recurrence. *Health Technol Assess* 2010;14:1-103.
32 Pletcher MJ, Pignone M. Evaluating the clinical utility of a biomarker: a review of methods for estimating health impact. *Circulation* 2011;123:1116-24.
33 Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;56:119-27.
34 Deeks JJ. Assessing outcomes following tests. In: Price CP, Christenson EH, eds. Evidence-based laboratory medicine: principles, practice and outcomes. 2nd ed. AACC Press; 2007;95-111.

Accepted: 30 November 2011

Cite this as: *BMJ* 2012;344:e686

Tables

Table 1| Attributes of the test-treatment pathway that affect patient health

Pathway component and mechanism	Definition
(1) Diagnostic test delivered	
Timing of test	Speed with which a test is performed within the management strategy
Feasibility	Completion of test process. Reasons for non-completion are: patient acceptability (patient's refusal to have test), test was contraindicated (clinical reason not to administer test), and technical failure (ability of diagnostic equipment to produce data)
Test process	Patients' interaction with test procedure, potentially causing physical or psychological harms or benefits
(2) Test result produced	
Interpretability	Degree to which test data can be used to inform a diagnostic classification
Accuracy	Ability of a test to distinguish between patients who have disease and those who do not
Timing of results	Speed with which test results are available
(3) Diagnosis made	
Timing of diagnosis	Speed with which a diagnostic decision is made
Diagnostic yield	Degree to which the test contributes to a patient diagnosis in any form, including: provision of a definitive diagnosis, confirmation of a suspected diagnosis, ruling out a working diagnosis, and distinguishing between alternative diagnoses with different treatment implications. Diagnostic yield is different from accuracy because it also incorporates any other information used by a doctor to make a diagnosis (such as previous test results)
Diagnostic confidence	Degree of confidence that doctors and patients have in the validity or applicability of a test result
(4) Management decided	
Therapeutic yield	Degree to which diagnostic decisions affect treatment plans
Therapeutic confidence	Certainty with which doctors and patients pursue a course of treatment
(5) Treatment implemented	
Timing of treatment	Speed with which patients receive treatment
Treatment efficacy	Ability of the treatment intervention to improve patient outcomes
Adherence	Extent to which patients participate in the management plan, as advised by their doctor, to attain therapeutic goal

Table 2| Checklist to determine clinically important differences between test-treatment pathways of new and existing diagnostic test strategies

Test-treatment pathway		Questions	Yes/no	Notes	Outcome to capture difference
Component	**Mechanism**				
(1) Test delivery	**Timing of test**				
	Time to test delivery	Do the strategies provide testing within comparable timeframes (or does one strategy deliver a diagnostic test considerably earlier than the other)?			
	Feasibility				
	Acceptability	Is the new test likely to be as acceptable to patients as the existing test (or does one test cause increased discomfort, for example)?			
	Clinical contraindications	Is the new test likely to be suitable in similar proportions of the relevant patient group (or will the new test be contraindicated in more or fewer patients than the existing test)?			
	Technical failure rates	Do the two tests produce similar proportions of failed procedures (or does the process of one test tend to fail more frequently than the other)?			
	Test process				
	Procedural harms or benefits	Are the two tests similar in how they affect patients during their application, physically or psychologically (or is one test more intrusive than the other or has a higher procedural morbidity than the other)?			
	Placebo effect	Does the new strategy give patients a similar perspective on being investigated (or could the new strategy encourage patients as to the thoroughness of their investigation)?			
(2) Test result	**Interpretability**				
	Ease of interpretation	Do the two processes produce similar frequencies of clearly interpretable test results (or once completed successfully, does one test tend to produce a higher frequency of indeterminate or unreadable results than the other)?			
	Accuracy				
	Accuracy	Do the tests correctly identify the target condition in the same number of patients (or does one test correctly identify a higher proportion of patients with disease or without disease than the other)?			
	Timing of results				
	Time to produce a result	Is the speed with which results are processed similar between tests (or does the new test have a shorter turnaround time between testing and production of results than the existing test)?			
(3) Diagnostic decision	**Timing of diagnosis**				
	Speed of diagnosis	Do the strategies produce diagnoses in comparable timeframes (or do patients given one test receive a diagnosis more quickly than patients given the other test)?			
	Diagnostic yield				
	Diagnoses made	Do the tests contribute to patient diagnosis to similar degrees (or do the results of one test tend to be given more weight than the other)?			
	Diagnostic confidence				
	Doctors' confidence in diagnosis	Is the degree of confidence that doctors have in the validity or applicability of a test result similar to that of its comparator test (or does a new test provide greater reassurance to doctors, or are its results considered less reliable by doctors)?			
	Patients' confidence in diagnosis	Is the degree of confidence that a patient has in the diagnostic process, or the diagnosis itself, likely to vary between strategies (or does a new test provide greater or lesser reassurance to patients, owing to doctors' confidence, the testing experience, or understanding of test results)?			
(4) Treatment decision	**Therapeutic yield**				
	Treatment choices	Do the comparative tests contribute to the formulation of a management plan to similar degrees (or does one test lead to more patients receiving appropriate treatment than the other)?			
	Therapeutic confidence				
	Doctors' confidence in treatment choice	Do doctors have similar confidence in pursuing a treatment plan between intervention arms (or does a test improve treatment success)?			
	Patients' confidence in treatment choice	Do patients have similar confidence in treatment plans based on diagnostic testing (or does the new test improve patients' understanding of the choice in management)?			

Table 2 (continued)

Test-treatment pathway		Questions	Yes/no	Notes	Outcome to capture difference
Component	**Mechanism**				
(5) Treatment implementation	**Timing of treatment**				
	Time to treatment	Do the diagnostic strategies lead to patients receiving treatment within comparable timeframes (or do patients given one test receive treatment earlier than those given the other test)?			
	Treatment efficacy				
	Efficacy of treatment	Does use of the intervention in patients identified to have disease lead to improvements in patient outcomes (or is the intervention ineffective)?			
	Adherence				
	Adherence to treatment	Are patients as likely to adhere to treatment plans regardless of the test strategy used (or does one strategy lead to more refusals or poorer compliance with treatment)?			

Figures

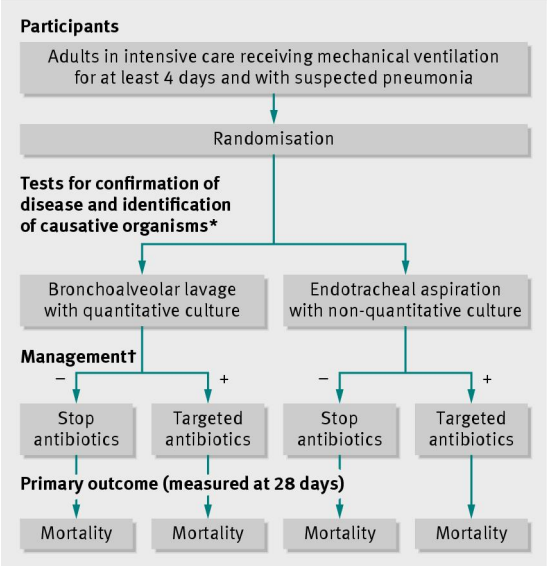

Fig 1 Design of a test-treatment randomised trial assessing whether bronchoalveolar lavage reduces the rate of death from ventilator associated pneumonia compared with endotracheal aspiration.[7] *All patients received broad spectrum antibiotics while waiting for test results. †In patients with confirmed pneumonia, antibiotics were adjusted using culture results and sensitivities; in test negative patients, antibiotics were discontinued

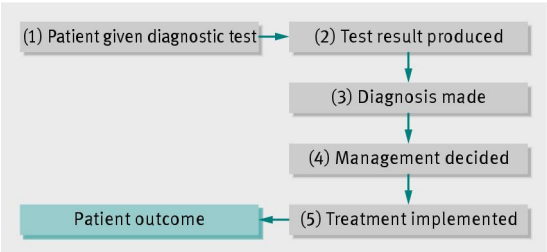

Fig 2 Simplified test-treatment pathway showing each component of a patient's management that can affect health outcomes[10]

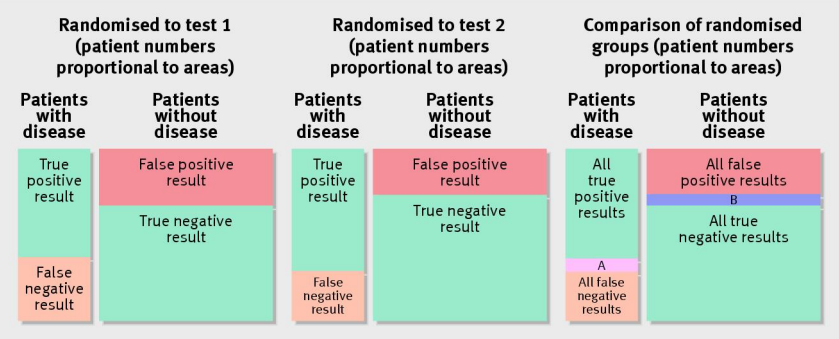

Fig 3 Sample size calculations for test-treatment randomised controlled trials. In randomised trials of interventions, all participants in a study group are allocated to receive the same intervention. In test-treatment trials, participants in each group receive a variety of interventions, depending on the test results and ensuing diagnosis. The magnitude of the observed treatment effect depends on the differences in proportions of patients who receive interventions appropriate to their condition in each group. This proportion would be expected to be quite small. The figure identifies those participants who contribute statistical power in a randomised trial comparing two tests (where the difference in outcome originates entirely from a difference in diagnostic accuracy). Test 2 has higher sensitivity than test 1 (difference shown in A). Test 2 also has higher specificity than test 1 (difference shown in B). Different widths of diseased and non-diseased columns indicate the prevalence of disease in the study sample. Only participants in A and B would have different test results if they received test 2 rather than test 1 and therefore the potential for different outcomes (all other participants in the study would have the same test result, irrespective of which test they were allocated to). Statistical power therefore depends on only the numbers of participants in A and B (particularly A); for example, if disease prevalence was 20%, and test 2 improved sensitivity by 20%, only 4% of the total sample size would fall in A[34]